Leveraging Deep Learning based Feature Extraction Techniques for Improved Audio Based Depression Detection

SHORT SYNOPSIS

Submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

by

Deepak Joshi
Registration No: MRUCSDN012K200470

Under the supervision of

Prof (Dr.) Manpreet Kaur
Professor

Dr. Harsh Bhasin
Assistant Professor, Bennett University

Department of Computer Science & Technology
School of Engineering

MANAV RACHNA UNIVERSITY

Year ***2025***

MANAV RACHNA UNIVERSITY

Date: 08/09/2025

DECLARATION BY THE SCHOLAR

I hereby declare that the short synopsis entitled **Leveraging Deep Learning based Feature Extraction Techniques for Improved Audio Based Depression Detection** being submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in the Department of **Computer Science & Technology** under the **School of Engineering** of Manav Rachna University, Faridabad during the academic year **2025-26**.

It is the work that is expected to be carried out under the supervision of **Dr. Manpreet Kaur** and co-supervision of **Dr. Harsh Bhasin**.

Further, it is declared that the short synopsis is within the plagiarism limit and copy of the same is attached as per the University norms.

Deepak Joshi
Registration Number - **MRUCSDN012K200470**

MANAV RACHNA UNIVERSITY
Declared as State Private University Vide Haryana UGC Act 26 of 2014

MANAV RACHNA UNIVERSITY

Activate Windows
Go to Settings to activate Windows.

Date: 08/09/2025

CERTIFICATE BY SUPERVISOR(S)

This is to certify that, the short synopsis entitled **Leveraging Deep Learning based Feature Extraction Techniques for Improved Audio Based Depression Detection** is an original piece of work done by research scholar **Deepak Joshi** under the supervision of **Dr. Manpreet Kaur** and co-supervision of **Dr. Harsh Bhasin** for the degree of Doctor of Philosophy in the **Department of Computer Science & Technology** under the **School of Engineering** of Manav Rachna University, Faridabad during the academic year **2025-26**.

Further, it is certified that the short synopsis is within the plagiarism limit and copy of the same is attached as per the University norms.

Manpreet Kaur

Dr. Harsh Bhasin

Supervisor

Prof (Dr.) Manpreet Kaur
Professor
Department of Computer Science & Technology
Manav Rachna University
Faridabad, Haryana

Co-Supervisor

Dr. Harsh Bhasin
Assistant Professor
Bennett University
Greater Noida

iii

ABSTRACT

In the entire world, Dispiritedness condition or commonly known as depression is evolving as one of the tough and serious physiological predicament. Across the globe a significant number of people are reportedly affected by this physiological disorder due to countless reasons. There are many conventional methods adopted by care practitioners to evaluate depression involving clinical assessment and scoring through questionnaires etc. However, that decision might be compromised due to the human bias and present mental condition of the interviewer along with other demographic factors. Therefore, this called out the depression detection in the automated mode where very less human interaction is required in detection. Therefore, in the next timeline various methods using Artificial Intelligence were developed for automatically depression detection. As of now there has been many research conducted in this domain using various modalities. In our study we will be using speech data for the detection of depression among other modalities (text, video, EEG etc.) through various Machine Learning/Deep Learning techniques. In this research the methodology is centered on remodeling and restructuring feature extraction techniques based upon deep learning for identifying subjects having strong sign of depression.

This research work will particularly exploit Distress analysis depression datasets publicly available to obtain the audio characteristics of the subjects and we will also arrange the handcrafted data by conducting health checkup camps in collaboration of psychiatrists and psychologist. Resultant output that produces data will undergo preprocessing and various Machine Learning/Deep Learning techniques will be applied to make the classification decision to identify the depressed patients and healthy persons.

Keywords: Artificial Intelligence, Audio, Depression dataset, Depressive Disorder, Psychologist

TABLE OF CONTENTS

S. No.	Description	Page No.
	Cover Page	i
	Declaration by Scholar	ii
	Certificate by Supervisor(s)	iii
	Abstract	iv
	Contents	v
1	**Introduction**	6-8
	1.1 Introduction to Depression	6
	1.2 Background of study	8
2	**Literature Review**	9-15
	2.1 Literature Review	9
	2.2 Gaps Identified	14
3	**Description of Broad Area**	15
4	**Objectives of the Research**	16
5	**Methodology to be adopted**	17-18
	5.1 Data Collection	17
	5.2 Data Preprocessing	17
	5.3 Feature Extraction	17
	5.4 Model	17
	5.5 Performance and Evaluation	18
7	**Expected Outcome of the Research**	18
8	**References**	18-22

1. Introduction

1.1 Introduction to Depression

In the present era, due to socio-economic development people's life have become very easy and therefore people have become very lethargic, self-centered and socially disassociated. This advancement has also pathway of many physical and mental health related issues and especially depression has popped up very rapidly. Presently depression is becoming one of the most challenging and leading psychosocial problem among other mental, psychological and emotional disorders which is affecting the major population of the world.

Depressed individual exhibits the behavior of persistent sadness, tendency of low feeling, lack of confidence, lost state of mind, social anxiety and less productivity at workplace and in extreme cases increased attempts of self-harming or suicidal thoughts [1].

The World health Organization (WHO) assessment indicated that around 280 million people across the globe suffered from depression that approximates nearly 3.8% of the entire population [2].Psychologists have categorized depression in three different levels based on individual's severity assessment namely mild or low-level depression which has extremely low impact in daily life, moderate or medium level depression which has considerable impact on individual's life and Severe or extreme level depression which makes lot of disturbance in performing day-to-day activities and suicidal tendencies [2].

Clinical psychologists and doctors have identified distinct patterns of depression episodes illustrated as Single episode which can be individual's first and only episode and the other one is Recurrent episode which implies that an individual have at least two episodes of depression in the past. Researchers have discovered various other medical conditions where depression is considered as a symptom. Depression disorder, anxiousness and stress are three most common problems that lead people to visit a psychologist. The feeling of being anxious overwhelmed and depressed are all so similar that there can be some overlap between depressive signs and other conditions. Nevertheless, there is subtle difference between all these conditions. While anxiety and stress are normal human experience and almost everyone might have confronted stressful and anxious situation and felt down from time to time, but

when these feelings persisting for at least more than two weeks, depression disorder manifests.

Depression can start and progress gradually however its effects can persist for longer duration or recurrent. For many individuals with depression don't even realize the symptoms or ignores it however in many cases a friend or family member notices the problem. Depression is not constrained to a specific class of people, all type of people that includes old, rich, poor and young etc. are affected by depression. Interestingly in many countries it has been observed that women are more susceptible to depressive disorder than men. Despite of having the massive impact and seriousness of this problem, there are many well identified, effective and structured treatments available. Line of treatment varies based on the intensity and type of depression which includes but not limited to regular counseling, therapeutic intervention, guided self-help, joining support group and medicines [3].

As per the survey results from WHO (World Health Organization) approximately more than 75 % of individuals globally having symptoms of depression residing in socio -economic low- and middle-income level countries doesn't get the treatment [2]. The biggest obstacle in obtaining the efficient treatment is expensive therapeutic treatment costs, lack of trained and well experienced mental health care professionals, inadequately resourced mental healthcare facilities and the social taboo associated with mental health issues [4].

There has been a cross-sectional study carried out in low level income country and the study revealed an important fact that approximately 39.16 % individuals of that country with symptoms of psychological disorders are often not correctly assessed and sometimes misdiagnosed with some other disease with a considerable amount of 54.72 % misdiagnosed with major depressive disorder [5].

It has been now established that if there are no adequate steps taken to treat depression then it will adversely impact individual's quality of life and also bothers people around him or connected to him.

Most of the manual techniques used for identifying depression are interview-style systems, assessment tools like the Hamilton Rating Scale for Depression and self-reporting methods like the PHQ-8 score and Beck Depression Inventory tool. Well organized Interviews by psychologists are intended to forecast the acuteness of symptoms and typical behaviors seen in the individuals suffering from depression [6], [7] [8].

Major challenges with these conventional approaches are that, in order to predict symptoms of depression, these methods require additional patient information, including background, history, and any information related to past traumatic situations. Additionally, these treatments require ongoing monitoring of the Individual's activities [9].

As the early diagnosis of depression using conventional manual is extremely time consuming as well as expensive and may have tendency to inaccurate diagnosis due to human bias therefore the strong need of automated techniques for identifying depression has emerged that could not only lessen the strain on current mental healthcare system but also assist medical professionals to make a suitable strategy to deal with this disease.

A thorough analysis of literature is conducted to examine the current methodologies and identify gaps in previous research publications.

1.2 Background of Study

In the recent years due to great advancement in Artificial Intelligence, researchers have shifted their focus on using AI techniques – ML and DL algorithms to detect the depression with increased accuracy without human bias [10] [11]. Assessing the mental state of an individual anyone can easily determine the depression. It has been observed by the researchers that a depressed person's voice is influenced by their prolonged psychoneurotic disorders. An individual with depression has change in its facial expressions, voice, pitch and tone and takes irregular pauses in the speech [12]. A very common sign of depression is delayed and inaccurate speech with sighs [13].

Due to these aspects, researchers have commenced to diagnose depression from individual's audio [14], video [15], text [16] and magnetic neuroimaging (MRI) [17] modalities and also with multimodal based techniques [18] and have attained promising results [19]. The outcome of those results implies that individual's verbal and visual processing can reveal social-behavioral patterns that mental health specialists may overlook, leading to improved depression detection and cost savings [19]. Some of the research revealed that video modality has superior accuracy than audio for diagnosing depressive disorder [20]. However, this modality poses concerns about patient privacy and data security. To safeguard patients' privacy, several databases do not make raw footage available [21] [22]. Apart from that

individual’s uneasiness in case of video recording camera also factors in giving inaccurate results thus makes it lesser reliable.

Moreover, diagnosis of depressive disorder with merely MRI alone doesn’t fit into providing an efficient and authentic diagnosis. Therefore, the efficacy of depression identification can be improved by combining both MRI and other modality [17].That said audio recordings are easier to capture huge amounts of data in a pleasant and safe atmosphere, unlike visual recordings. In addition to that audio recordings are more reliable and promising and can preserve patients' privacy. As a result, we concentrate on how to accurately identify depression through speech [19].

The general procedure of depression detection using traditional ML approaches comprises preprocessing data, followed by feature extraction and using relevant techniques to minimize the dimension, and ultimately classification algorithms to identify depression [15]. Some of the researchers employed conventional approaches for feature extraction i.e. hand-crafted feature extraction whereas others used DL for automatic feature extraction [23]. Considering the above factors and previous researches in this area, we will be carrying out our study with the aim of optimal and highly accurate prediction of depression using audio data.

2. Literature Review

2.1 Literature Review

From this time forward a profuse amount of exploration has been accomplished on automated ways of detecting depression using audio, video and other modes. A thorough analysis of literature is conducted to examine the current methodologies and identify gaps in previous research publications.

Das et al. [24] formulated an multimodal framework based approach to determine the depressive disorder by amalgamating the most discriminative features pulled out from the audio based spectrogram as well Mel-Frequency Cepstral Coefficients (MFCC) [25]. Audio features derived from above methods are undergo trained using neural network based classifier. In order to evaluate the performance of models this study made use of audio based

publicly available datasets DAIC-WOZ [22] and MODMA [26]. In addition to above, for emotion recognition RAVDASS [27] dataset has been utilized.

Yong et al. [19] developed a hierarchical depression diagnosis model that gives the guidance to predict the severity level of depression by employing the technique of orchestrating the various classifiers in advance than the repressor. This proposed work leveraged the pre trained models to extricated the features related to speaker recognition as well as emotion recognition from speech and thereafter both are fused together to improve the accuracy of determining prediction. The model performance has been tested in AVEC 2013 [28] and AVEC 2014 [29] database.

In an another study performed by Zhang et al. [18] a multimodal framework was proposed that works on audio, text and video modals. This work highlights the two important tasks, one is for detecting the depression whereas the other is to determine the severity of depressive symptoms. There are three different techniques applied to perform the aforementioned tasks. Pre-trained model sentence modelling for text along with Bidirectional Long Short term memory (BiLSTM), SVM has been used for audio and extreme gradient boosting has employed for video [18].

Tian et al. [14] proposed a decision model that performs the tasks of emotion recognition by pulling out the audio characteristics using OpenSmile followed with choosing the most discriminative features using feature selection techniques. Furthermore, the study proposes the usage of convolutional neural network for model testing and analysis.

In 2022 Miao et al. [30] introduced the feature fusion technique by combines the higher order spectral features consists from bicoherent and bispectral features and the conventional speech features with the classification weights. In order to evaluate the fused features both conventional methods which consist of support vector machine and KNN classification and deep Neural Network.

Rejaibi et al. [25] in 2022 conducted a research on determining the depression and its severity level using the MFCC based high and low level audio features and using the DAIC – WOZ dataset with an assumption that MFCC based features are good candidates for depressive disorder detection.

In 2019 Huang et al. [31] proposed the short term mood disorder detection using the speech responses of the individual where emotion profile was constructed by CNN and LSTM [32] based attention level mechanism.

The likelihood of individual's psychiatric illness has been explored in a study performed by kour et al. [33] in 2022 by using the textual data obtained from tweets. In this study a hybrid model that consist of CNN and bi-directional LSTM has been proposed.

An automated feature extraction has been proposed in a research performed by zhong et al. [34] in 2023 which aids in early detection of depressive symptoms during the clinical assessment of the patients and this method ensures the data relationship between constructed weighed graphs.

The Table 1 shows the details of the literature reviewed and the findings from it.

Title, Year	Model / Method	Modality	Major Findings
A Multimodal approach for Detection and Assessment of Depression Using Text, Audio and Video [35], 2024	An ensemble voting algorithm based fusion model that used multiple modalities for detection.	Text, Audio Video	This study has developed a multimodal fusion model with audio, text and video to diagnosing and assessment of depression [35]
A novel hybrid optimization algorithm for depression detection using MRI and speech signal [17], 2023	Multimodal fusion model using brain MRI image and Audio for diagnosing using Deep Maxout Network [17]	Audio, MRI	This research has used entirely two different modalities, MRI image and speech signal for discovering the depression in adolescents using MPI–Leipzig dataset [17].
Attention guided learnable time-domain filterbanks for speech depression detection [36], 2023	The model developed here is called DALF which is Attention based depression identification model [36].	Audio	Chinese depression dataset called the Neutral Reading-based Audio Corpus (NRAC), DAIC-woz [36].
A multimodal fusion model with multi-level attention mechanism for depression detection [37], 2023	The research study come up with the fusion model having muti-level attention mechanism [37].	Audio, Visual, Text	This paper has used seven kinds of data from various modalities to use the feature diversity and also focus on information redundancy [37].
A novel study for depression detecting using audio signals based on graph neural network [38], 2023	GNN (Graph Neural Network) [38].	Audio	This paper has proposed the model which explores the inter class variability and intra-class consistency of audios [38].
Smart voice recognition based on deep learning for	This study come up with the idea of creating deep	Audio	This study has used voice data obtained from the depressive

depression diagnosis [39], 2023	residual network based on deep learning model [39].		individuals who are taking therapy session and also the data from university students who are healthy and non- depressed [39].
Depression detection using cascaded attention based deep learning framework using speech data [40],2023	DL framework called DAttn_Conv 2D LSTM [40].	Audio	The study use deep learning based model to detect depression using DAIC-WOZ) and Sonde Health-free speech (SH2-FS) datasets [40].
Prediction of Depression Severity Based on the Prosodic and Semantic Feature with Bi-directional LSTM and Time Distributed CNN [41], 2023	Bidirectional Long Short-Term Memory Network (Bi-LSTM) followed by a Time-distributed Convolutional Neural Network (T-CNN) [41].	Audio, Text	This research article use Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ) dataset and create attention based multimodal to detect and asses the severity of depression) [41].
A Model of Normality Inspired Deep Learning Framework for Depression Relapse Prediction Using Audiovisual Data [42], 2022	This research defines Model of Normality based framework [42].	Audio, Video	This study illustrates reoccurrence of the depressive symptoms by the proximity of the audio and visual features patterns of an individual after a symptomless duration to the patterns of depressed subjects [42].
Enhanced depression detection from speech using Quantum Whale Optimization Algorithm for feature selection [43] , 2022	This study has used various classification methods which are LDA, KNN Classifier, SVM and Logistic Regression [43].	Audio	This study describes a special type of feature selection algorithm having two phases and it chooses the optimum subset of the discriminant and distinct speech features for diagnosing depression [43].
MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech [25], 2022	The MFCC-based RNN presents three LSTM layers followed by two fully connected layers [25].	Audio	This study proposes a Recurrent Neural Network-based framework for depression detection and its severity level prediction from speech [25].
What You Say or How You Say It? Depression Detection Through Joint Modeling of Linguistic and Acoustic Aspects of Speech [44], 2022	Bi-LSTM and Gated Multimodal Units [44].	Audio, Text	This study is based on networks designed for sequence modeling and multimodal analysis methodologies [44].
Fusing features of speech for depression classification based on higher-order spectral analysis [30], 2022	In this research, for conventional ML models– Support Vector Machine and K Nearest Neighbor for deep	Audio	In this study, two models are proposed traditional machine learning models and deep learning models for depression recognition by using speech-related features [30].

	learning model, CNN model has been used [30].		
Audio based depression detection using Convolutional Autoencoder [23], 2022	CNN Autoencoder (CNN AE) [23].	Audio	This study focus on learning most relevant and distinct feature from the raw sequential data and have used cluster based technique to address sample imbalance problem [23].
Multi-modal depression detection based on emotional audio and evaluation text [45], 2021	1.Audio Model-Temporal Convolutional Network 2. Text model -Introduced the model "one-hot Transformer" 3. Multimodal fusion Att-TCN Transformer to fuse the two modes at the decision level [45].	Audio, Text	This study describes multimodal fusion method and word vector features for diagnosis depression by using deep learning [45].
A hierarchical depression detection model based on vocal and emotional cues, [19] 2021	New hierarchical depression detection model [19].	Audio	This study details about extracting speaker recognition and speech emotion recognition features using pre-trained models, and combine the two deep speech features to take advantage of the complementary information between the vocal and emotional differences of speakers [19].
An End-to-End Model for Detection and Assessment of Depression Levels using Speech [46], 2020	Spectrogram based CNN and end to end CNN models are implemented [46].	Audio	In this study depression are detected using audio samples using Mel - spectrogram based CNN and E2E traditional neural networks [46].
Automated speech-based screening of depression using deep convolutional neural networks [47], 2019	Pre-trained CNN architectures that used a residual learning framework ResNet-18, 34, 50, 101, 152 [47].	Audio	This study describes multipart interactive training network using spectrograms [47].
Attention-based convolutional neural network and long short-term memory for short-term detection of mood disorders based on elicited speech responses [31], 2019	The model being used here is CNN with an attention mechanism [31].	Audio	This study describes mood disorder and labeled Multimedia Human–Machine Communication emotion database for training of model and evaluation for mood disorder detection. [31]

Table 1: Literature Review

2.2 Gaps Identified

Based on the literature survey, various gaps are identified as follows.

1. There have been several research work published where the sample size for research is extremely small enough that results in model overfitting and therefore model gives surprised and inaccurate results. This overfitting issue dents the overall performance of the model.
2. For some of the studies, video data has been used for prediction of depression which is relatively difficult to assess and can potentially flair up individual's personal identity issues. Owing to this, relatively lesser population gives the consent to use this data for the detection.
3. Few of the studies have used the data obtained through non-surgical medical imaging techniques such as brain MRI; however potentially ethical concerns are likely to be raised regarding the use of such sensitive medical data. Due to inappropriate trained technicians or imperfect machines there might be possibility of wrong parameters from cortex area of brain images are captures which leads to inaccurate results. Moreover if such data has to be manually collected it requires enormous amount of resource chain and approvals from the respective committees who takes care of community mental health.
4. Some other studies indicated the Imbalance between depressed and non-depressed samples in the dataset that can cause biased model performance and may affect the generalizability of the results.
5. In few of the studies the authors have solely relied on social media text data for predicting depression which alone is not appropriate for depression analysis as this works best in conjunction with other modalities.
6. Few papers have baselined their study only on the specific population and language – i.e. Women / Chinese speakers which makes the study limited to a particular demography and not scaled up for wider population.

Rationale behind choosing Audio Modality for study

Choosing audio data for diagnosing depression has its own advantages as the manifestation of depressive symptoms occurs first in individual's speech and language and can be easily identified by anyone without doing any clinical test. Speech analysis is considered as the best biomarkers for depression detection due to its simple, easily collection and readily availability. Speech patterns and features of the depressed person is easily distinguishable than in other modes. Unlike video mode where individual's identity can be revealed and non-desire to give consent to use samples for research, audio is safe and has no such issues.

Owing to this our research is purely focused covering all the gaps mentioned above and use speech data for the diagnosis of depression that don't compromise the individual identity and ambition to create ML/DL models in a way that don't overfit or underfit with the provided data and don't give surprised results.

Problem Statement

The present research work hypothesizes to devise audio data depression detection method with expected quality closer to an ideal conventional depression detection by experienced healthcare professional. An exemplary automated AI system has the characteristics of being paramount accuracy, less redundant and more intelligible for the person using it. Therefore, the problem statement for the present study is formulated as below:

To develop an automatic depression identification system using deep learning which accurately identifies the presence or absence of depressive symptoms of an individual through analysis of their speech samples.

3. Description of the Broad Area

It is a general perception that health always comes first into mind. On the top of that healthy mental well-being has paramount importance in everyone's life. There are many inventions made in the healthcare domain and due to the advancement individual's life span has been increased and as a result many diseases are now being cured by various medical practices such as therapies, medicines and surgical interventions. However, the mental and

psychological disease detection and assessment are still done by the healthcare professionals in the conventional way like questionnaires and interviews which can potentially give incorrect assessment if the healthcare person is not trained or biased or affected by mental and emotional conditions. Therefore, this problem calls out the automated way of detecting the mental health related issue. Depression or depressive disorder emerged as very common psychological problem which affects the person's life in every aspect and also bothers people connected with the person. Therefore, depression detection turns out to be a hot topic in these days.

Research Area: -

Mental Health & Machine Learning

Research Title: -

Leveraging Deep Learning based Feature Extraction Techniques for Improved Audio Based Depression Detection

4. Objectives of Research

The current research work will explore audio data for depression detection. The objectives of the current proposed research work are to:

1. Conduct a comprehensive review of conventional Machine Learning pipeline for detecting depression from audio data.

2. Study existing state of the Art methods and analyzing the performance for depression detection using audio data.

3. Develop and evaluate Deep Learning based methods for depression detection on audio data.

4. Demonstrate the improvement in the performance and efficiency of the detection technique by comparing it with existing similar techniques.

5. Methodology to be adopted

As mentioned before, the study will be aimed at diagnosing the depression using the speech data obtained from publicly available and homegrown data collected from various surveys. Below steps explains the methodology adopted in the current work.

- Data Collection
- Data Preprocessing
- Feature Selection/Extraction
- Model Training
- Performance Evaluation

5.1 Data Collection

For this study, popular publicly depression datasets will be used. In addition to that homegrown dataset is planned to be created by organizing surveys and mental health camps with the collaboration with Healthcare professionals.

5.2 Data Preprocessing

In the proposed study dataset will be partitioned in training, development, and test sets while preserving the overall speaker diversity. The speech data will be preprocessed to remove the additional noise and interviewer's voice in order to improve the audio quality interviewee.

5.3 Feature Extraction

Various Feature extraction techniques will be applying on the processed data to get the meaningful features for detecting depression.

5.4 Model Training

After meaningful features will be extracted from speech samples, it will be passed to various Deep Learning models including pretrained models with self-attention mechanism and transformers. Furthermore, fusion techniques will be applied to

combine the features derived from extraction methods to obtain the accurate prediction.

5.5 Performance Evaluation

The performance metrics i.e. recall, F1 score etc. will be used to evaluate the performance of the model and further it will be compared with other state- of the art methods.

6. Expected outcome of the research

In overall, the expected outcome is the formation of the depression diagnostic system using ML/DL techniques. The diagnostic system thus created will be enable to classify the depressed patients among the healthy population with improved accuracy. The proposed system will not only identify the depression but it will be able to evaluate the episode/stage of the depression. To further evaluate the correctness of the system, it will be plan to testify with the clinical healthcare professional. The long-term aim is to scale up this system for commercial purpose by embedding this into the PCB using IOT technique and use it as a preliminary diagnostic tool in those places where there is a shortage of trained phycologists.

7. References

[1] V. Higuera, "Everything You Need to Know About Depression (Major Depressive Disorder)," 27 10 2023. [Online]. Available: https://www.healthline.com/health/depression#prevention.

[2] "Depressive disorder (depression)," 31 March 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[3] "Overview - Depression in adults," [Online]. Available: https://www.nhs.uk/mental-health/conditions/depression-in-adults/overview/.

[4] A. K. S. M. A. e. a. Kaur, "Mental health related stigma, service provision and utilization in Northern India: situational analysis," *Int J Ment Health Syst,* p. 17, 2023.

[5] G. D. S. y. Z. e. a. Ayano, "Misdiagnosis, detection rate, and associated factors of severe psychiatric disorders in specialized psychiatry centers in Ethiopia," *Ann Gen Psychiatry 20*, 2021.

[6] H. Y. Q. C. D. H. Y. W. X. Ma, "DepAudioNet: an efficient deep model for audio based depression classification, in: AVEC 2016," *AVEC '16: Proceedings of the 6th*

International Workshop on Audio/Visual Emotion Challenge, p. 35–42, 2016.

[7] F. G.-S. N. K. T. F. J. G. B. Arroll, "Effect of the addition of a 'help' question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study," *Br. Med. J.*, p. 884–886, 2005.

[8] e. a. J.F. Cohn, "Detecting depression from facial actions and vocal prosody," in *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, Amsterdam, 2009.

[9] F. &. B. A. &. N. E. &. S. S. &. K. J. Hoenig, "Automatic modelling of depressed speech: relevant features and relevance of gender," in *Proceedings of the Annual Conference of the International Speech Communication Association,*.

[10] K. C. Q. C. a. L. Y. M. Niu, "HCAG: A Hierarchical Context-Aware Graph Attention Model for Depression Detection," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics*, Toronto, 2021.

[11] H. Z. Q. B. Y. Z. L. W. G. L. Zhijun Dai, "Improving depression prediction using a novel feature selection algorithm coupled with context-aware analysis," *Journal of Affective Disorders*, vol. 295, pp. 1040-1048, 2021.

[12] N. M. D. C. Vandana, "A hybrid model for depression detection using deep learning," *Measurement: Sensors*, vol. 25, February 2023.

[13] H. D. a. W. L. G. Lam, "Context-aware Deep Learning for Multi-modal Depression Detection," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics*, Brighton, 2019.

[14] Z. Z. X. J. Han Tian, "Deep learning for Depression Recognition from Speech," *Mobile Networks and Applications*, 2023.

[15] N. K. A. S. M. S. R. P. B. Dr Harsh Bhasin, "Kullback-Leibler Divergence Based Method for Depression Diagnosis Using Video Data," in *Computing Communication and Networking Technologies (ICCCNT)*, 2023.

[16] S. A. T. L. P. C. Kuan-Chieh Lu, "Depression detection via conversation turn classification," *Multimedia Tools and Applications*, vol. 82, 2023.

[17] P. M. V. W. Minakshee Patil, "A novel hybrid optimization algorithm for depression detection using MRI and speech signal," *Biomedical Signal Processing and Control*, 2023.

[18] K. M. C. Wei Zhang, "A Multimodal Approach for Detection and Assessment of Depression," *International Human Phenome Institutes (Shanghai)*, 2023.

[19] X. Y. Yizhuo Dong, "A hierarchical depression detection model based on vocal and emotional cues," *Neurocomputing,*, vol. 441, pp. 279-290, 2021.

[20] D. Y. A. N. J. N. B. H. T. Q. R. Williamson, "Tracking depression severity from audio and video based on speech articulatory coordination," *Computer Speech Language*, p. 40–56, 2019.

[21] J. G. B. S. F. R. D. L. M. T. T. S. S. G. S. R. C. M. P. M. Valstar, "AVEC 2016 - Depression, mood, and emotion recognition workshop and challenge," in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016.

[22] B. S. M. V. J. G. R. C. S. S. S. M. N. C. M. S. M. P. F. Ringeval, "Avec 2017: Real-life depression, and affect recognition workshop and challenge," in *Proceedings of the 7th*

Annual Workshop on Audio/Visual Emotion Challenge, 2017.

[23] B. N. M. N. R. P. E. Sara Sardari, "Audio based depression detection using Convolutional Autoencoder," *Expert Systems with Applications*, vol. 189, 2022.

[24] Arnab Kumar Das , Ruchira Naskar, "A deep learning model for depression detection based on MFCC and CNN generated spectrogram features," *Biomedical Signal Processing and Control*, vol. 90, 2024.

[25] A. K. F. M. S. A. O. Emna Rejaibi, "MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech," *Biomedical Signal Processing and Control*, vol. 71, 2022.

[26] Z. Y. Y. G. S. S. N. L. F. T. H. X. J. L. Z. Y. X. L. e. a. H. Cai, "A multi-modal open dataset for mental-disorder analysis," *Scientific Data 9*, 19 April 2022.

[27] F. A. R. Steven R. Livingstone, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," 16 May 2018.

[28] T. F. Q. B. S. H. R. H. B. Y. a. D. D. M. James R. Williamson, "Vocal biomarkers of depression based on motor incoordination," in *In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge (AVEC '13)*, 2013.

[29] T. F. Q. B. S. H. G. C. a. D. D. M. James R. Williamson, "Vocal and Facial Biomarkers of Depression based on Motor Incoordination and Timing," in *In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC '14)*, 2014.

[30] Y. L. M. W. Y. L. I. N. J. H. G. Xiaolin Miao, "Fusing features of speech for depression classification based on higher-order spectral analysis," *Speech Communication*, pp. Pages 46-56, 2022.

[31] C.-H. W. M.-H. S. Kun-Yi Huang, "Attention-based convolutional neural network and long short-term memory for short-term detection of mood disorders based on elicited speech responses," *Pattern Recognition*, vol. 88, pp. 668-678, 2019.

[32] J. S. Sepp Hochreiter, "Long Short-Term Memory," *Neural Computation (1997) 9 (8)*, p. 1735–1780, 1997.

[33] Harnain Kour,Manoj K. Gupta , "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," *Multimedia Tools and Applications*, vol. 81, p. 23649–23685, 2022.

[34] Jitao Zhong,Wenyan Du ,Lu Zhang,Hong Peng,Bin Hu, "Feature extraction based on sparse graphs embedding for automatic depression detection," *Biomedical Signal Processing and Control*, 2023.

[35] W. M. K. &. C. J. Zhang, "A Multimodal Approach for Detection and Assessment of Depression Using Text, Audio and Video," *Phenomics*, vol. 4, pp. 234-249, 2024.

[36] J. L. C. Z. W. K. L. W. Z. Wenju Yang, "Attention guided learnable time-domain filterbanks for speech depression detection," *Neural Networks*, vol. 165, pp. 135-149, 2023.

[37] S. P. Y. L.-C. H. L. Ming Fang, "A multimodal fusion model with multi-level attention mechanism for depression detection," *Biomedical Signal Processing and Control*, vol. 82, 2023.

[38] M. J. L. G. Y. X. D. Chenjian Sun, "A novel study for depression detecting using audio

signals based on graph neural network," *Biomedical Signal Processing and Control,* vol. 88, 2024.

[39] N. S. Sukit Suparatpinyo, "Smart voice recognition based on deep learning for depression diagnosis," 2022.

[40] G. A. A. P. Sachi Gupta, "Depression detection using cascaded attention based deep learning framework using speech data," *Multimedia Tools and Applications,* 2023.

[41] W. Zhang, D. B. Wang, A. Li, R. Jiao and Y. Zhu, "Prediction of Depression Severity Based on the Prosodic and Semantic Features With Bidirectional LSTM and Time Distributed CNN," in *IEEE,* 2022.

[42] A.-O. Z. M. M. Alice Othmani, "A Model of Normality Inspired Deep Learning Framework for Depression Relapse Prediction Using Audiovisual Data," *Computer Methods and Programs in Biomedicine,* vol. 226, 2022.

[43] S. R. R. A. Baljeet Kaur, "Enhanced depression detection from speech using Quantum Whale Optimization Algorithm for feature selection," *Computers in Biology and Medicine,* vol. 150, 2022.

[44] A. E. V. Nujud Aloshban, "What You Say or How You Say It? Depression Detection Through Joint Modeling of Linguistic and Acoustic Aspects of Speech," *Cognitive Computation,* pp. 1585-1598, 2022.

[45] Y. Y. W. L. L. Z. G. F. Jiayu Ye, "Multi-modal depression detection based on emotional audio and evaluation text," *Journal of Affective Disorders,* vol. 295, pp. 904-913, 2021.

[46] S. L. N.S. Srimadhur, "An End-to-End Model for Detection and Assessment of Depression Levels using Speech," in *Procedia Computer Science,* 2020.

[47] K. W. I. K. Karol Chlasta, "Automated speech-based screening of depression using deep convolutional neural networks," *Procedia Computer Science,* vol. 164, pp. 618-628, 2019.

[48] Karol Chlasta, Krzysztof Wołk, Izabela Krejtz,, "Automated speech-based screening of depression using deep convolutional neural networks," *Procedia Computer Science,* vol. 164, pp. 618-628, 2019.

[49] "Depressive disorder (depression)," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[50] N. K. A. S. M. S. a. R. P. B. H. Bhasin, "Kullback-Leibler Divergence Based Method for Depression Diagnosis Using Video Data," in *14th International Conference on Computing Communication and Networking Technologies (ICCCNT),* Delhi, 2023.

[51] Vandana, Nikhil Marriwala,Deepti Chaudhary, "A hybrid model for depression detection using deep learning," *Measurement: Sensors,* 2023.

[52] Yurui Xu,Hang Su, Guijin Ma, Xiaorui Liu, "A novel dual-modal emotion recognition algorithm with fusing hybrid features of audio signal and speech context," *Complex & Intelligent Systems,* 2023.

[53] Minakshee Patil,Prachi Mukharji, Vijay Wadhai, "A novel hybrid optimization algorithm for depression detection using MRI and speech signal," *Biomedical Signal Processing and Control,* 2023.

[54] Sara Sardari, Bahresh Nakisa, Mohammed Naim Rastgoo,Peter Eklund, "Audio based depression detection using Convolutional Autoencoder," *Expert Systems With*

Applications , vol. 189, 2022.

[55] Shamim Mobram,Mansour Vali, "Depression detection based on linear and nonlinear speech features in I-vector/SVDA framework," *Computers in Biology and Medicine,* vol. 149, 2022.

[56] Yujuan Xing,Zhenyu Liu,Qiongqiong Chen,Gang Li,Zhijie Ding,Lei Feng,Bin Hu, "Depression recognition base on acoustic speech model of Multi-task emotional stimulus," *Biomedical Signal Processing and Control,* vol. 85, 2023.

[57] Shanaya Rathod, Narsimha Pinninti,Muhammed Irfan,Paul Gorczynski,Pranay Rathod,Lina Gega,and Farooq Naeem, "Mental Health Service Provision in Low- and Middle-Income Countries," *Health Serv Insights.,* vol. 10, 2017.

[58] "Multi-feature deep supervised voiceprint adversarial network for depression recognition from speech," *Biomedical Signal Processing and Control,* vol. 89, 2024.

[59] Mingyue Niu,Jianhua Tao,Bin Liu,Jian,Zheng Lian, "Multimodal Spatiotemporal Representation for Automatic Depression Level Detection," *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING.*

[60] Kaining Mao,Wei Zhang,Deborah Baofeng Wang, Ang Li,Rongqi Jiao,Yanhui Zhu,Bin Wu,Tiansheng Zheng,Lei Qian,, "Prediction of Depression Severity Based on the Prosodic and Semantic Features With Bidirectional LSTM and Time Distributed CNN," *IEEE Transactions on Affective Computing,* vol. 14, no. 3, pp. 2251 - 2265, 2023.

[61] Yazheng Di, Jingying Wang,Weidong Li, Tingshao Zhu, "Using i-vectors from voice features to identify major depressive disorder," *Journal of Affective Disorders,* vol. 288, pp. 161-166, 2021.

MANAV RACHNA UNIVERSITY

Swami Vivekananda Library and Resource Centre
Manav Rachna University

SIMILARITY VERIFICATION CERTIFICATE

Title: Leveraging Deep Learning based Feature Extraction Techniques For Improved Audio based Depression Detection

Researcher: Deepak Joshi

Registration Number: MRUCSDNO12K2004

Supervisor: Prof. Dr. Manpreet Kaur

Department: Computer Science & Technolo

This is to certify that the above mentioned Short Synopsis/Thesis was scanned for similarity index through Drillbit Software on 04/09/25 with submission ID 4315454 and the similarity index was found to be 1 %. The complete report running into 19 Pages, 5978 Word count with File Size 1270 KB is being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 15/09/25

15/9/2025

The complete report of the above Short Synopsis/Thesis has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/Thesis may be considered for further processing

Deepak Joshi
Research Scholar

Manpreet Kaur
Research Supervisor
8/9/25

Excluding matches less than 14 words

DrillBit

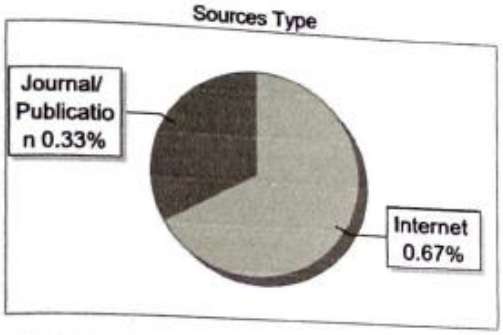
The Report is Generated by DrillBit Plagiarism Detection Software

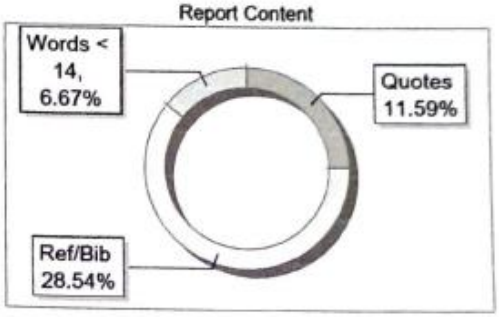
Submission Information

Author Name	Deepak Joshi
Title	Leveraging Deep Learning based Feature Extraction Techniques for Improved Audio Based Depression Detection
Paper/Submission ID	4315454
Submitted by	netrapal.library@mru.edu.in
Submission Date	2025-09-04 13:36:01
Total Pages, Total Words	19, 5978
Document type	Synopsis

Result Information

Similarity **1 %**





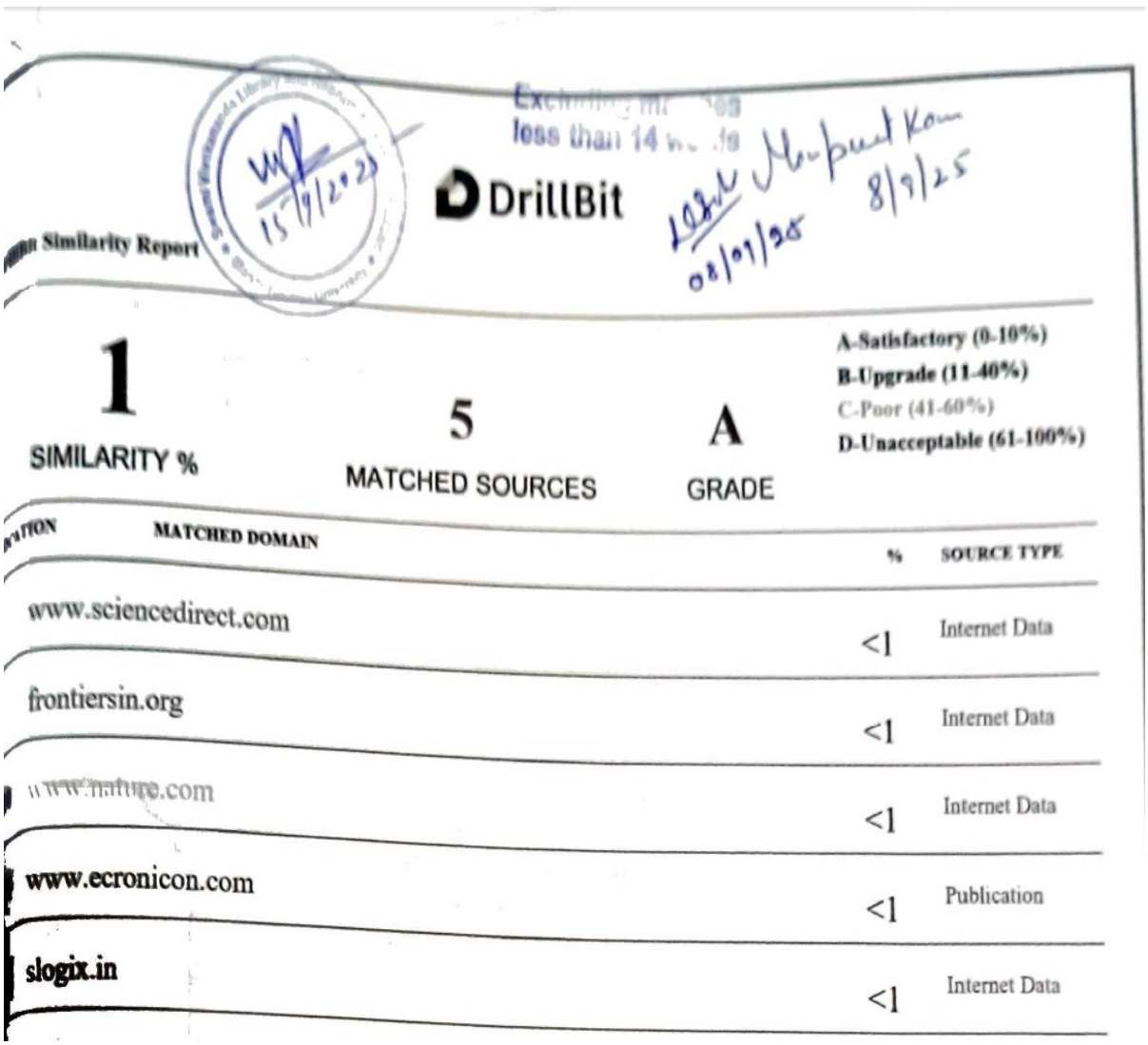
Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File



DrillBit

Similarity Report

less than 14 words

8/9/25

1	5	A
SIMILARITY %	MATCHED SOURCES	GRADE

A-Satisfactory (0-10%)
B-Upgrade (11-40%)
C-Poor (41-60%)
D-Unacceptable (61-100%)

MATCHED DOMAIN	%	SOURCE TYPE
www.sciencedirect.com	<1	Internet Data
frontiersin.org	<1	Internet Data
www.nature.com	<1	Internet Data
www.ecronicon.com	<1	Publication
slogix.in	<1	Internet Data

Leveraging Deep Learning based Feature Extraction Techniques for Improved Audio Based Depression Detection

SHORT SYNOPSIS

by

Deepak Joshi

Registration no: MRUCSDN012K200470

Under the supervision of

Prof (Dr.) Manpreet Kaur
Professor

&

Dr. Harsh Bhasin
Assistant Professor

MANAV RACHNA UNIVERSITY

Year
2025

ABSTRACT

In the entire world, Dispiritedness condition or commonly known as depression is evolving as one of the tough and serious physiological predicament. Across the globe a significant number of people are reportedly affected by this physiological disorder due to countless reasons. There are many conventional methods adopted by care practitioners to evaluate depression involving clinical assessment and scoring through questionnaires etc. However, that decision might be compromised due to the human bias and present mental condition of the interviewer along with other demographic factors. Therefore, this called out the depression detection in the automated mode where very less human interaction is required in detection. Therefore, in the next timeline various methods using Artificial Intelligence were developed for automatically depression detection. As of now there has been many research conducted in this domain using various modalities. In our study we will be using speech data for the detection of depression among other modalities (text, video, EEG etc.) through various Machine Learning/Deep Learning techniques. In this research the methodology is centered on remodeling and restructuring feature extraction techniques based upon deep learning for identifying subjects having strong sign of depression.

This research work will particularly exploit Distress analysis depression datasets publicly available to obtain the audio characteristics of the subjects and we will also arrange the handcrafted data by conducting health checkup camps in collaboration of psychiatrists and psychologist. Resultant output that produces data will undergo preprocessing and various Machine Learning/Deep Learning techniques will be applied to make the classification decision to identify the depressed patients and healthy persons.

Keywords: Artificial Intelligence, Audio, Depression dataset, Depressive Disorder, Psychologist

15/9/2025

DrillBit

08/09/25

The Report is Generated by DrillBit AI Content Detection Software

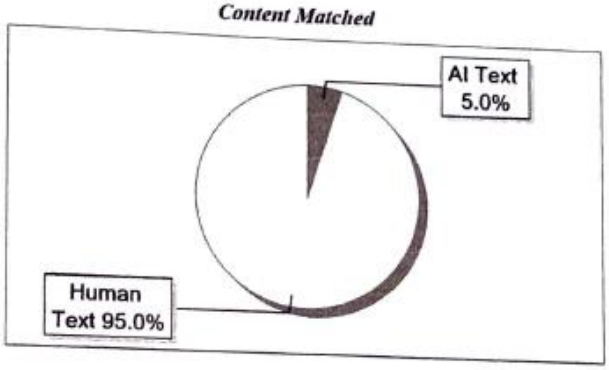
Submission Information

Author Name	Deepak Joshi
Title	Leveraging Deep Learning based Feature Extracti..
Paper/Submission ID	4315454
Submitted By	netrapal.library@mru.edu.in
Submission Date	2025-09-04 13:36:01
Total Pages	19
Document type	Synopsis

Result Information

AI Text: **5 %**

Content Matched



Disclaimer:

- **The content detection** system employed here is powered by artificial intelligence (AI) technology.
- **Its not always accurate** and only help to author identify text that might be prepared by a AI tool.
- **It is designed to** assist in identifying & moderating content that may violate community guidelines/legal regulations, it may not be perfect.